<Bioinformatics Center>Mathematical Bioinformatics

Bioinformatics Center
– Mathematical Bioinformatics –

http://www.bic.kyoto-u.ac.jp/takutsu/index.html

Prof
AKUTSU, Tatsuya
(D Eng)

Assoc Prof
TAMURA, Takeyuki
(D Inf)

SPD (JSPS)
LIN, Chun-Yu
(Ph D)

Students

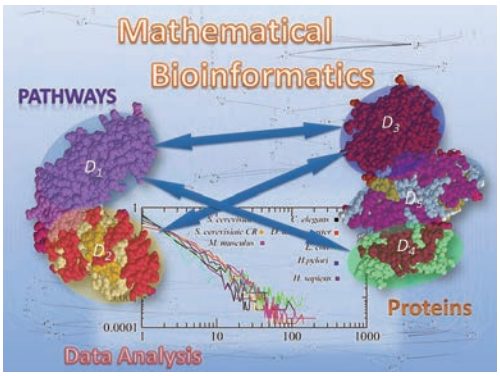
BAO, Yu (D3)
LIU, Pengyu (D2)
YU, Coleman (D1)
TAKAGI, Motoshige (D1)
LI, Ruiming (M2)
CAO, Yu (M1)
MARUTA, Kunpei (M1)
PI, Wenya (M1)
WANG, Feiqi (RS)

Scope of Research

Due to rapid progress of genome sequencing technology, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are currently studying the following topics: systems biology, scale-free networks, protein structure prediction, inference of biological networks, chemo-informatics, and discrete and stochastic methods for bioinformatics.

KEYWORDS

Complex Networks
Boolean Networks
Neural Networks
Chemical Graphs
Protein Informatics



Selected Publications

Akutsu, T.; Jansson, J.; Takasu, A.; Tamura, T., On the Parameterized Complexity of Associative and Commutative Unification, *Theor. Comput. Sci.*, **660,** 57-74 (2017).

Cheng, X.; Tamura, T.; Ching, W. K.; Akutsu, T., Discrimination of Singleton and Periodic Attractors in Boolean Networks, *Automatica*, **84,** 205-213 (2017).

Kato, Y.; Mori, T.; Sato, S.; Maegawa, S.; Hosokawa, H.; Akutsu, T., An Accessibility-Incorporated Method for Accurate Prediction of RNA-RNA Interactions from Sequence Data, *Bioinformatics*, **33,** 202-209 (2017).

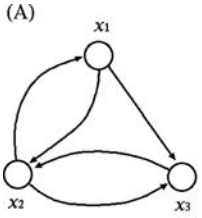
Ishitsuka, M.; Akutsu, T.; Nacher, J. C., Critical Controllability Analysis of Directed Biological Networks Using Efficient Graph Reduction, *Sci. Rep.*, **7,** [14361-1]-[14361-10] (2017).

Exact Identification of the Structure of a Probabilistic Boolean Network from Samples

Various kinds of mathematical models have been utilized for understanding dynamical behavior of biological systems. Among them, the Boolean network (BN) is a simple but well-studied discrete model, especially for modeling genetic regulatory networks. In a BN, each node takes a Boolean value, 0 or 1, at each time step, and the states of all nodes are updated synchronously according to Boolean functions assigned to nodes, where each node corresponds to a gene, and 1 and 0 mean that genes are active and inactive, respectively. Since a BN is a deterministic model and thus cannot cope with such effects as noise and uncertainty, several probabilistic extensions of a BN have been proposed and studied. Among them, the probabilistic Boolean network (PBN) has been well-studied. Different from a BN, multiple Boolean functions can be assigned to each node in a PBN and one function is randomly selected at each time step according to the prescribed probability distribution (Figure 1).

For both BNs and PBNs, it is important to infer the network model from such data as gene expression time series data and many methods have been proposed. However, almost no theoretical studies have been done on inference of a PBN from sample data. Since it is quite difficult to exactly determine the probabilities from samples, we focus on determining only the structure (graph structure + Boolean functions) of a PBN and study the number of samples required for uniquely determining the structure. We show via theoretical analysis and computer simulation that the structure of a PBN can be exactly identified with high probability from a relatively small number of samples for interesting classes of PBNs of bounded indegree (i.e., the number of edges per node is bounded by a constant). On the other hand, we also show that there exist classes of PBNs for which it is impossible to uniquely determine the structure of a PBN from samples. We are also performing detailed theoretical analyses with focusing on an important subclass of PBNs which consist of Boolean threshold functions.

(A)



(B)

	Boolean Function	Probability
$f_1^{(1)}$	$x_1(t+1) = x_2(t)$	0.9
$f_2^{(1)}$	$x_1(t+1) = \overline{x_2(t)}$	0.1
$f_1^{(2)}$	$x_2(t+1) = x_1(t) \wedge \overline{x_3(t)}$	0.3
$f_2^{(2)}$	$x_2(t+1) = x_3(t)$	0.7
$f_1^{(3)}$	$x_3(t+1) = x_1(t) \wedge \overline{x_2(t)}$	1.0

Figure 1. Example of a probabilistic Boolean network (PBN). (A) Graph structure of PBN. (B) Boolean functions and corresponding probabilities assigned to each node.

Observability of Singleton and Periodic Attractors in Boolean Networks

Knowing internal states of complex systems is important for diagnosing various kinds of biological, artificial, and social systems. Especially, it is important to identify a small set of variables so that we can reconstruct the system's complete internal state at any given time step from time-series data of these variables. In such a case, the system is called observable. This observability problem has been well-studied for linear systems. However, biological systems contain non-linear elements to which existing theory/methods cannot be effectively applied. Furthermore, existing studies suggest that a large number of variables/nodes are required to observe the whole state of certain kinds of non-linear biological systems.

In order to cope with this difficult situation, we do not intend to identify the whole state instead focus on identification of statically and periodically stable states (attractors) because attractors are often considered to correspond to cell types. In this study, we adopt a Boolean network (BN) as a non-linear model of biological systems and consider the problem of identifying a minimum set of sensor nodes to discriminate static and periodic attractors using the BN (Figure 2), which might be useful to identify cell types. We prove that one node is not necessarily enough but two nodes are always enough to discriminate two periodic attractors by making use of the Chinese remainder theorem. Based on this, we develop an algorithm to determine the minimum number of nodes to discriminate all given attractors. The results of computational experiments suggest that attractors (corresponding to cell types) in realistic BN models can be discriminated by observing the states of only a small number of nodes.

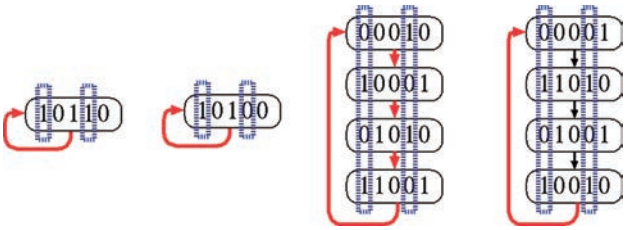


Figure 2. These four attractors (two statically stable and two periodically stable states) can be discriminated by observing time series data of two nodes (shown by blue dotted curves), where four attractors correspond to four different cell types.